

Article

An Entropy-Based Investigation into Bivariate Drought Analysis in China

Dongdong Zuo ^{1,2,†,‡}, Wei Hou ^{3,‡} and Jingguo Hu ^{1,2,*}¹ Institute of Mathematics and Science, Yangzhou University, Yangzhou 225002, China; dongdz_vow@163.com² Institute of Physics Science and Technology, Yangzhou University, Yangzhou 225002, China³ Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081, China; houwei@cma.gov.cn

* Correspondence: hujingguo_yzu@163.com

† Current address: Institute of Mathematics and Science, Yangzhou University, Yangzhou 225002, China.

‡ These authors contributed equally to this work.

Received: 17 June 2017; Accepted: 20 August 2017; Published: 23 August 2017

Abstract: Because of the high correlation between random variables of drought duration and severity, their joint distribution is difficult to obtain by traditional mathematical methods. However, the copula method has proved to be a useful tool for analyzing the frequency of drought duration and severity. Most studies have used different marginal distribution functions to fit the drought duration and severity distributions. This requires a great deal of contrast analysis, and sometimes two or more distributions fit the data well. Based on entropy theory, however, a unified probability distribution function is derived which reduces complex contrast analysis and improves the filtering distribution function. Based on monthly precipitation data at 162 stations in China for 1961–2015, the monthly standardized precipitation index was calculated and used to extract drought duration and severity. Then the entropy distribution was used to fit the distributions of drought duration and severity, and to establish the correspondence between them. The probabilities of the interval and return periods were then determined using the copula method. An analysis of the discrepancy between the conventional and entropy-based methods indicated that the entropy distribution showed a better fit than conventional methods for drought duration distribution, although no obvious difference was found in drought severity distribution. The entropy-based results were more consistent with the empirical data, whereas conventional methods showed apparent deviation in some drought types. Hence, the entropy-based method is proposed as an alternative method of deriving the marginal distributions of drought duration and severity, and for analyzing the interval probability and return period in China.

Keywords: drought; entropy; copula; China

1. Introduction

Of the many natural hazards, droughts are a common and major hazard, not only for their devastating impact on regional agriculture, but also for their far-reaching impact in an increasingly globalized world [1,2]. The large spatial coverage and long duration of droughts are their main characteristics. According to Wilhite [3], droughts cause global damage costing tens of billions of dollars. Overall, they affect more people than any other form of devastating climate-related hazard. Drought is a serious problem in China, as demonstrated by the annual proportion of crop-damaged area due to drought. According to data from the National Bureau of Statistics in China, in the period 2004–2013 the average annual area of crop damage due to drought accounted for 50% of all areas affected by meteorological disasters (<http://data.stats.gov.cn/easyquery.htm?cn=C01>).

To date, many drought indices have been proposed for quantifying, monitoring and analyzing droughts [4,5], but the drought problem is a complex one involving a great many factors [6–12]. There is no unified index because of the complexity of the physical processes involved; however, many studies have shown that the intensity, duration and spatial extent are the main characteristics of droughts [13,14]. Although the combined probability of these characteristics is important in drought analysis, drought variables display different marginal distributions. Despite a certain degree of correlation, it is difficult to calculate their joint function. For this reason, only limited analyses of the combined duration and severity of droughts have so far been conducted.

The copula method was introduced in recent years to analyze droughts. A copula function is a multivariate descriptive model used in probability analysis, originally applied in economics, then in hydrology and meteorology. The detailed theoretical background and descriptions for the use of copulas can be acquired from the books of Joe [15] and Nelsen [16]. Some recent works include, for example, Shiau [17] who constructed a joint drought duration and severity distribution using bivariate copulas in southern Taiwan; Zhang [18] employed a copula function to analyze drought risk, indicating a high risk in central and north-western Yunnan Province; and Tosunoglu and Can [19] used a two-dimensional copula function to analyze meteorological droughts in Turkey. Similar analyses can be found in [20–23].

However, most of these studies used different marginal distribution functions to fit the drought duration and severity distributions. For instance, Reddy and Singh [24] used gamma, lognormal, Weibull and exponential distributions to fit drought duration and severity, and showed that a lognormal distribution produced the best fit for drought severity, and an exponential distribution produced the best fit for drought duration. Ganguli and Reddy [25] used a bivariate copula to assess the drought risk in Gujarat, India, adopting a normal kernel, quadratic kernel, gamma, lognormal, exponential, Gumbel and Weibull distributions to fit the drought duration and severity; these indicated that the exponential and normal kernel distributions respectively produced the best fit for drought duration and severity. Similar analyses can be found in [19] and [23]. Hence, a large amount of contrast analysis must be conducted to select an optimal distribution and, in some cases, two or more distributions fit the data well. Selecting the most appropriate marginal distribution is a tedious process but, based on the principle of maximum entropy [26,27] developed from Shannon entropy [28], a unified distribution formulation was derived and the results produced a good fit in practice [29–33]. Since this method has been more often applied in the analysis of hydrological drought than meteorological drought, a discussion of the results of entropy-based and conventional methods is necessary and meaningful.

The major object of this study is to analyze meteorological drought in China adopting entropy methods, and to discuss the discrepancy between these and conventional methods. This paper is organized as follows. Section 2 describes the data used in this study. Section 3 defines drought characteristics and distributions, drought classification, entropy distribution and the copula method. The results are given in Section 4. Finally, discussion and main conclusions of this study are summarized in Section 5.

2. Data

Monthly precipitation data from 194 national meteorological observation stations throughout mainland China was obtained from the China meteorological data service center (<http://data.cma.cn/en>). However, for various reasons (missing data, differences in start times, and to ensure that more historical observations were included), 162 stations were selected for the period from 1961 to 2015 (Figure 1).

As shown in Figure 1, due to the lack of meteorological stations in Tibet, the results for Tibet are masked. The mean annual precipitation distribution exhibits an obvious pattern, decreasing from south-eastern to north-western China. The maximum mean annual precipitation was above 1800 mm and the minimum was less than 300 mm. This great difference is due to the complex topography of China, with mountain ridges, river basins, plateaus, hills and plains, as well as major climate differences.

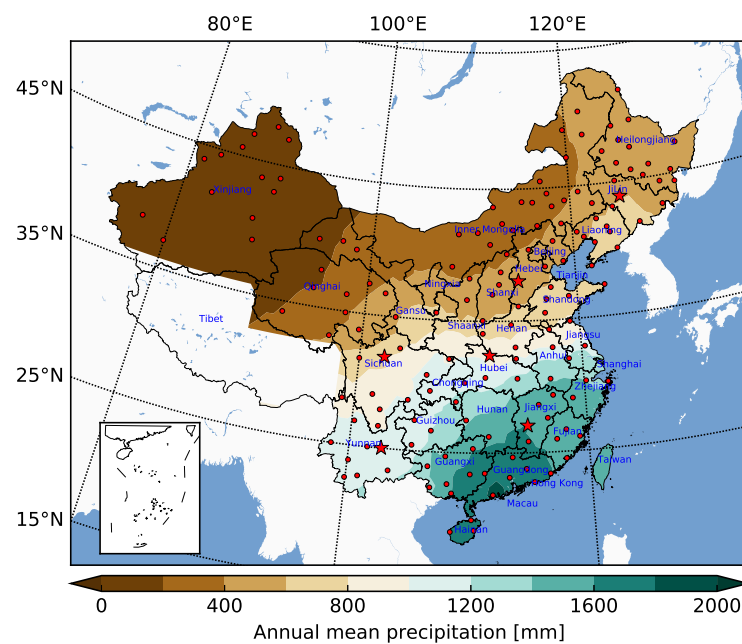


Figure 1. Locations of meteorological observation stations and contours of mean annual precipitation in China. Stars represent selected stations: Changchun (station No. 54161) in Jilin; Shijiazhuang (station No. 53698) in Hebei; Laohekou (station No. 57265) in Hubei; Maerkang (station No. 56172) in Sichuan; Kunming (station No. 56778) in Yunnan; and Jianxian (station No. 57799) in Jiangxi.

3. Methodology

3.1. Standardized Precipitation Index (SPI)

The SPI is a simple and widely used drought index which was developed by McKee [34]. The number of SPI denotes the standard deviations that the observed value would deviate from the long-term mean, for a normally distributed random variable. Since precipitation is not normally distributed, a transformation is first applied so that the transformed precipitation values follow a normal distribution. The SPI expresses droughts on different time scales, such as 1, 3, 6 or 12 months, and so on. A detailed description of the calculation steps may be found in Guttman [35]. The program may be downloaded at <http://drought.unl.edu/MonitoringTools/DownloadableSPIProgram.aspx>. The fact that the SPI is calculated from precipitation data makes it relatively easy to evaluate, and thus it is ideal for areas where data has not been collected extensively. Because of these advantages, it is widely used to investigate drought characteristics; see [36–39].

3.2. Definition of Drought Characteristics and Distributions

Duration and severity are the two main properties of drought event, which were extracted from the SPI (1 month scale) in this article, as shown in Figure 2. Drought is defined as a continuous period when the SPI is below zero [17]; therefore, the drought duration is equal to the number of months when the SPI is continuously below zero. Drought severity is defined as the cumulative values of the SPI within the drought duration, expressed as

$$s_i = - \sum_{i=1}^{d_i} SPI_i \quad (1)$$

where d_i is the drought duration and s_i is the drought severity.

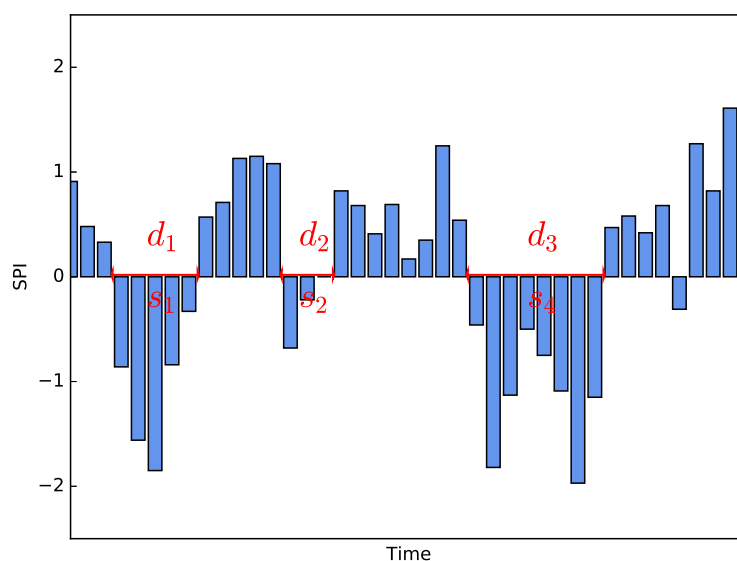


Figure 2. Definitions of drought characteristics.

Previous studies [17,20,40] have suggested that drought duration and severity follow exponential and gamma distributions, respectively. The cumulative probability functions of the exponential distribution and gamma distribution are given by

$$F_D(D \leq d) = 1 - e^{-\lambda d} \quad (2)$$

where d is the drought duration and λ is a parameter, estimated by $\hat{\lambda} = 1/\bar{d}$; $F_D(D \leq d)$ is the cumulative distribution function for duration which refers to the probability that drought duration is equal to or less than d ; and

$$F_S(S \leq s) = \int_0^s \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{s}{\beta}} ds \quad (3)$$

where s is the drought severity; Γ is the gamma function; and α and β are shape and scale parameters, respectively, estimated by $\hat{\alpha} = \frac{1}{4A}(1 + \sqrt{1 + \frac{4A}{3}})$, $\hat{\beta} = \frac{\bar{s}}{\hat{\alpha}}$, where $A = \ln \bar{s} - \overline{\ln s}$; $F_S(S \leq s)$ is the cumulative distribution function for severity which denotes to the probability that drought severity is equal to or less than s .

In practice, the time is usually classified into a specific time scale (e.g., month, season, half-year, one year etc.). In the present study, the drought duration (D) was divided into four classifications (Table 1). Based on the empirical cumulative probability of drought severity (S), the percentage of drought severity was calculated to have the same probability as the classification of drought duration indicated. These percentages were also used to divide the severity into four types (Table 1).

Table 1. Classifications of drought duration and severity.

Duration (Month)	Severity	Probability	Classification
$0 < D \leq 1$	$0 < S \leq 0.85$	0.39	1
$1 < D \leq 3$	$0.85 < S \leq 2.27$	0.39	2
$3 < D \leq 6$	$2.27 < S \leq 4.29$	0.17	3
$6 < D$	$4.29 < S$	0.05	4

The probability 0.05 is often used as the percentage point defining extreme event. As Table 1 shows, the probability of $6 < D$ and $4.29 < S$ is 0.05; therefore, $D = 6$ and $S = 4.29$ indicates the

bounds of an extreme event. By combining the classifications of duration and severity in Table 1, 16 types of drought events are obtained: drought types where $0 < D \leq 1$ and $0 < S \leq 0.85$, $0 < D \leq 1$ and $0.85 < S \leq 2.27$, etc.

3.3. Entropy-Based Distribution

3.3.1. Shannon Entropy

The term entropy [28] describes the uncertainty, disorder, dispersion or diversification of a system. For a continuous random variable X with probability density function (PDF) $f(x)$ defined on the interval $[a, b]$, Shannon entropy H is defined as

$$H = - \int_a^b f(x) \ln f(x) dx \quad (4)$$

3.3.2. Univariate Entropy

Univariate distribution was derived from the principle of maximum entropy proposed by Jaynes [26,27] where the probability density function (PDF) maximizes the entropy, subject to given constraints. The general constraints are given by

$$\int_a^b f(x) dx = 1 \quad (5)$$

$$\int_a^b g_i(x) f(x) dx = \overline{g_i(x)}, i = 1, 2, \dots, k \quad (6)$$

where the constraint in Equation (5) ensures that the integration of the PDF over the interval $[a, b]$ is unity. Equation (6) describes the other constraints, which can be selected or specified functions with respect to the properties of interest [32]; $\overline{g_i(x)}$ is the expected value of the i -th function $g_i(x)$, and k is the number of constraints. In this study the power function was selected as the $g_i(x)$ function, taking the parameter estimation to be a gamma distribution; a logarithmic function was also used. Hence Equation (6) may be rewritten as

$$\int_a^b x^i f(x) dx = \overline{x^i}, i = 1, 2, \dots, k \quad (7)$$

$$\int_a^b \ln x f(x) dx = \overline{\ln x} \quad (8)$$

In accordance with the principle of maximum entropy, the entropy-based PDF for the univariate case is derived by maximizing the entropy defined in Equation (4) using Lagrange multipliers. The Lagrange function L is given by [41]:

$$L = - \int_a^b f(x) \ln f(x) dx - \sum_{i=0}^k \lambda_i \left[\int_a^b g_i(x) f(x) dx - \overline{g_i} \right] \quad (9)$$

where λ_i are the Lagrange multipliers.

Differentiating L with respect to f and setting the derivative to zero, the entropy-based PDF can be obtained as [41,42]

$$f(x) = \exp \left[- \sum_{i=0}^m \lambda_i g_i(x) \right] \quad (10)$$

With the constraints in Equation (7) (the first to third powers of x), the entropy-based PDF of drought duration is expressed by

$$f(d) = \exp[-\lambda_0 - \lambda_1 d - \lambda_2 d^2 - \lambda_3 d^3] \quad (11)$$

Similarly, the entropy-based PDF of drought severity based on the constraints in Equation (7) (the first and second powers of x) and Equation (8) is expressed by

$$f(s) = \exp[-\lambda_0 - \lambda_1 s - \lambda_2 s^2 - \lambda_3 \ln s] \quad (12)$$

The cumulative distributions for the random variables (drought duration and severity) are then obtained by integrating each PDF.

The Lagrange multipliers are obtained by minimizing the convex function [41,43]

$$Y = \lambda_0 + \sum_{i=0}^4 \lambda_i \overline{g_i(x)} = \ln \int_a^b \exp[-\sum_{i=1}^4 \lambda_i g_i(x)] dx + \sum_{i=0}^4 \lambda_i \overline{g_i(x)} \quad (13)$$

The Newton-Raphson method can be used to minimize the convex function and obtain the Lagrange multiplier $\{\lambda_i, i = 0, 1, 2, 3\}$ [29,31] but here the `scipy.optimize.minimize` function in Python (<https://www.python.org>) software was used.

3.4. Copula

3.4.1. Definitions

Based only on the marginal distributions of the variables, a copula function gives the multivariate distribution function [16]. Sklar [44] first introduced the theoretical basis of a copula: if random variables x, y follow the arbitrary marginal distribution functions $F_X(x), F_Y(y)$, respectively, then a copula $C(\bullet)$ exists that combines the marginal distribution functions to give the joint distribution function $F_{X,Y}(x, y)$ [16]

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)) \quad (14)$$

3.4.2. Archimedean Copulas

There are many families of copulas, including elliptical (normal and t), Archimedean (Clayton, Gumbel and Frank), extreme value (Gumbel, Husler-Reiss, Galambos, Tawn and t -EV) and others (Plackett and Farlie-Gumbel-Morgenstern) [45]. Archimedean copulas are most commonly used for drought applications, and were employed in this study, to establish the joint probability expressed as follows [16].

Clayton copula:

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad (15)$$

where $C_\theta(u, v)$ is the joint distribution; u and v are marginal distributions; θ is a parameter of copula function which can be estimated by the relationship between the Kendal coefficient and the parameter θ expressed as $\tau = \frac{\theta}{\theta+2}, \theta \in [0, \infty)$.

Frank copula:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right), \theta \neq 0 \quad (16)$$

where the relationship between the Kendal coefficient and the parameter θ is $\tau = 1 - \frac{4}{\theta} \left(\frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt - 1\right)$, $\theta \neq 0$.

Gumbel-Hougaard copula:

$$C_{\theta}(u, v) = e^{-((- \ln u)^{\theta} + (- \ln v)^{\theta})^{\frac{1}{\theta}}} \quad (17)$$

where the relationship between the Kendal coefficient and the parameter θ is $\tau = 1 - \theta^{-1}, \theta \in [0, \infty)$.

3.4.3. Empirical Copulas

When a sufficiently large sample size is available, empirical copulas can be used to construct a non-parametric joint empirical probability distribution [16]. Unfortunately, the sample size is small in many cases, so an analysis using empirical copulas is not realistic; however, they are often used for copula selection. For the bivariate case, the empirical copula of the observed data (u_i, v_i) is

$$C_e(u_i, v_i) = \frac{1}{n} \sum_{i=1}^n I\left(\frac{D_i}{n+1} \leq u_i, \frac{S_i}{n+1} \leq v_i\right), \quad (18)$$

where n is the sample size; $I(A)$ is the indicator function, which is equal to 1 if A is true and 0 if A is false; and D_i and S_i respectively represent the drought duration and severity rank statistics obtained from the sample.

3.4.4. Copula Selection

The root mean squared error (RMSE), the Akaike information criterion (AIC) [46,47], and the Kolmogorov-Smirnov (KS) D_n statistic [48] were used to select the copula function with the best fit. These are expressed as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_{\theta}(u_i, v_i) - C_e(u_i, v_i))^2} \quad (19)$$

$$AIC = n \ln MSE + 2m \quad (20)$$

$$D_n = \sup |C_{\theta}(u_i, v_i) - C_e(u_i, v_i)| \quad (21)$$

where n is the sample size; C_{θ} is the computed value of the copula parametric; C_e is the observed value of the probability obtained from the empirical copula; MSE is the mean square error; m is the number of independently adjusted parameters; and sup is the supremum of the set of distances. This model is most efficient when RMSE, AIC and D_n take minimum values.

3.4.5. Interval Probability and Return Period

The occurring probability and return period of drought event are important information for water resource management, where they can be obtained from the derived copula-based joint drought duration and severity distribution. The joint probability and interval probability are given by

$$P_{d,s} = P(D \leq d, S \leq s) = C(F_D(d), F_S(s)) \quad (22)$$

$$\bar{P}_{d_1, d_2, s_1, s_2} = \bar{P}(d_1 < D \leq d_2, s_1 < S \leq s_2) = P_{d_2, s_2} - P_{d_1, s_2} - P_{d_2, s_1} + P_{d_1, s_1} \quad (23)$$

where $P_{d,s}$ is the cumulative joint probability which refers to the probability that both the drought duration and severity are equal to or less than certain thresholds, and $\bar{P}_{d_1, d_2, s_1, s_2}$ is the interval probability which denotes the occurring probability of a type of drought event denoted in Table 1.

The return period for droughts with a duration or severity equal to or greater than a certain value can be found in Shiau and Shen [17,20], but here the focus was on the return period of drought type indicated in Table 1 [49]. This is expressed as

$$T = \frac{n_{years}}{N \times \bar{P}_{d_1, d_2, s_1, s_2}} \quad (24)$$

where n_{years} is the range of years; N is the number size of drought samples extracted from the SPI sequence; and $\bar{P}_{d_1, d_2, s_1, s_2}$ is the interval probability.

4. Results

4.1. Drought Characteristic and Distribution Test

As indicated in Figure 2, drought is defined as a continuous period when the SPI is below zero. Hence, the drought events (containing two variables called drought duration and severity) were extracted from the time sequence of 1-month SPI at each station, and the number of drought events was counted which shows that the number of drought events is generally larger than 130. Taking the selected stations (Changchun and Maerkang in Figure 1) for example, Figure 3 shows the time sequences of drought duration and severity. The results of Kendall coefficient indicate that there is an ordinal association between drought duration and severity (0.52 in Changchun and 0.57 in Maerkang), and the Kendal coefficients pass the tau test at the 0.05 significance level. The Kendall coefficient in other stations was also calculated, which ranges from 0.45 to 0.73.

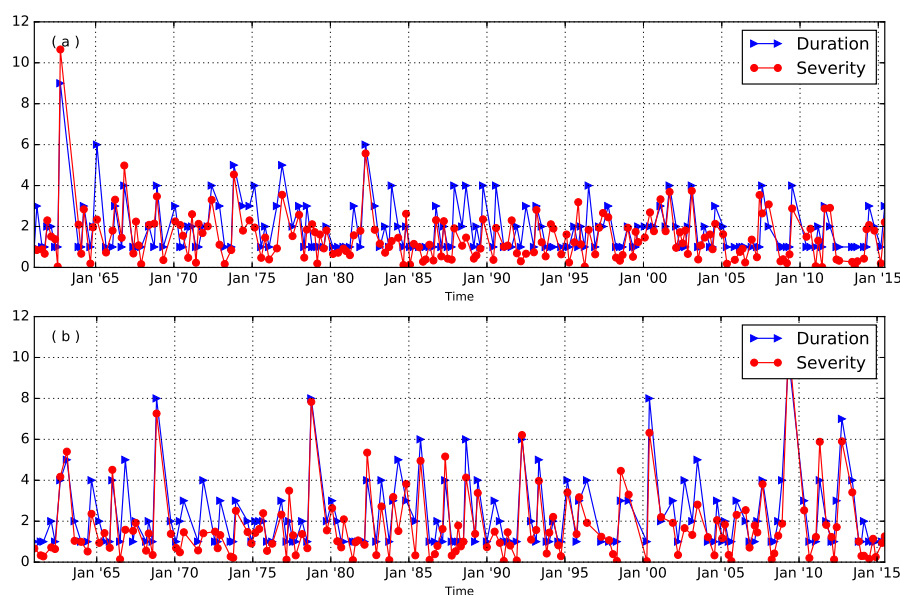


Figure 3. The time sequences of drought duration and severity: (a) Changchun (station No. 54161); (b) Maerkang (station No. 56172).

Based on the extracted drought duration and severity samples, the mean and standard deviation of the two drought characteristics were obtained (Figure 4). Figure 4a shows that the mean duration in most regions is close to two months, and the mean duration in southern, south-western and northern China is relatively large. Figure 4b shows a similar distribution, but Figure 4c clearly differs from Figure 4d, indicating greater variation of drought duration in south-western than in northern China, whereas drought severity shows a larger variation in south-western and south-eastern China. As indicated in other studies [50–56], droughts occur frequently in south-western, central and northern China. An analysis of the standard deviation of the two drought characteristics indicates that variation

in the duration of droughts is the main factor affecting northern China, but the south-west is affected by the combined effects of both the duration and severity of droughts.

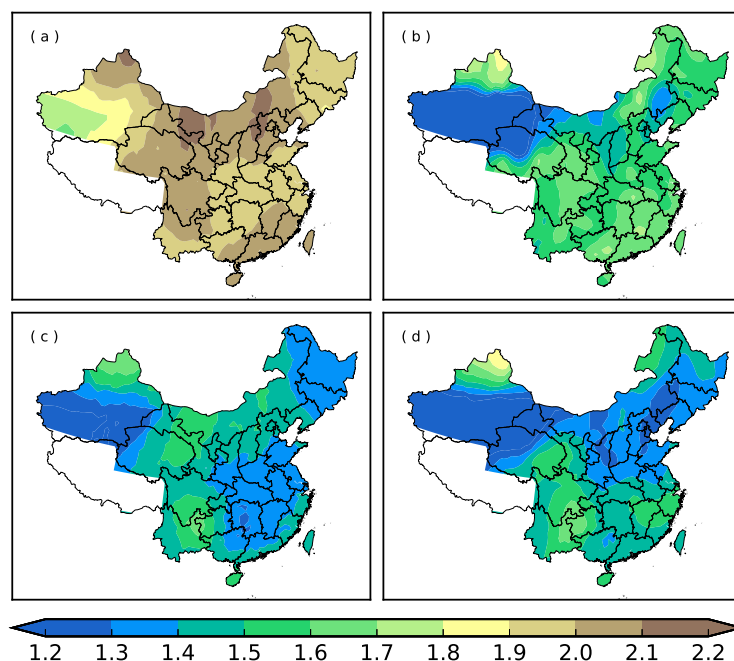


Figure 4. Distribution of drought characteristics: (a) mean drought duration; (b) mean drought severity; (c) standard deviation of drought duration; and (d) standard deviation of drought severity.

Generally, exponential distribution (Equation (2)) is used to model drought duration, while gamma distribution (Equation (3)) fits drought severity. However, in this study the entropy-based distribution was used to model the marginal distributions of both variables (see Equations (11) and (12)), with the empirical distribution being used for comparison. For the evenly selected stations in Figure 1, the theoretical cumulative distribution, entropy-based cumulative distribution and empirical cumulative distribution are shown in Figures 5 and 6. Overall, the theoretical distribution fitted the empirical distribution well, especially for drought severity, but discrepancies between the theoretical and empirical distributions were still evident. Figure 5 shows that the entropy-based distribution was more consistent with the data than the exponential distribution, which overestimated the probability when $D < 2$ and underestimated it when $D > 2$. However, as Figure 6 shows, the discrepancy is not distinct, and the distributions produced by both methods were reasonably consistent with the empirical distribution. The results of a KS goodness-of-fit test at all stations are shown in Table 2.

The KS results for all stations (Table 2) illustrate that drought duration is distributed exponentially, and drought severity obeys a gamma distribution. It is also interesting to note that these two variables also conformed to the entropy-based distributions. The root mean squared error (RMSE) and KS statistic D_n between theoretical (entropy) and empirical methods were calculated for each station. The mean of these two variables in Table 2 indicates that the entropy-based method produced a closer fit, especially for drought duration.

Since most studies in the literature have shown [18,19,22,23] that the Archimedean copulas are the most widely used for drought analysis, they were also used in this study to build a correspondence between the drought variables. Due to the limited number of samples, however, Equation (18) was used only to select the optimal copula function; it was not used to calculate the empirical interval probability and return period. However, for the analysis of the conventional process, the empirical marginal distribution and copula function were used to calculate the empirical interval probability

and return period that were to be adopted as the reference object (called as semiempirical method in our paper).

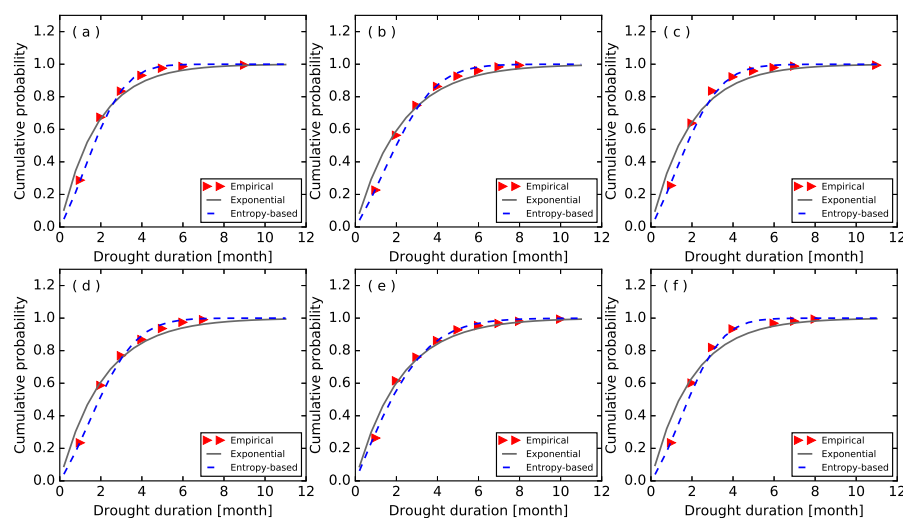


Figure 5. Cumulative distribution of drought duration at selected stations: (a) Changchun (station No. 54161); (b) Shijiazhuang (station No. 53698); (c) Laohekou (station No. 57265); (d) Maerkang (station No. 56172); (e) Kunming (station No. 56778); and (f) Jianxian (station No. 57799).

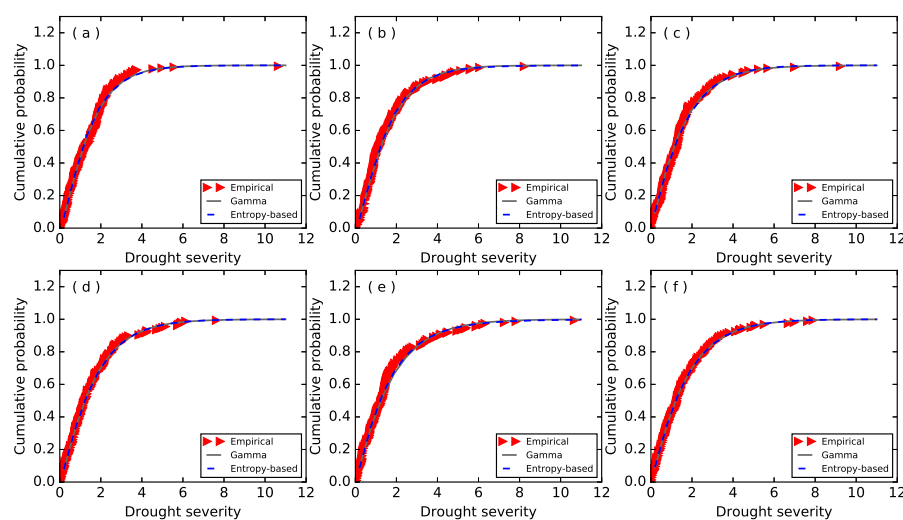


Figure 6. Cumulative distribution of drought severity at selected stations (a–f) as in Figure 5.

Table 2. Passing rate in KS test ($\alpha = 0.05$) and mean $RMSE$ and KS statistic D_n .

Variables	Duration		Severity	
	Exponential	Entropy-Based	Gamma	Entropy-Based
Proportion	100%	100%	100%	100%
D_n	0.141	0.063	0.051	0.046
$RMSE$	0.058	0.029	0.021	0.018

Figure 7 shows a theoretical estimation obtained using the Archimedean copula plotted versus the empirical copula. The 45° line indicates the consistency between theoretical joint probability and empirical joint probability. The results in Figure 7 shows that three copulas all show a good fit effect. In order to select optimal copula function, the RMSE, AIC, and D_n methods were used to select the copula function with the best fit (Table 3). As Table 3 indicates, the Gumbel copula was found to produce the best fit to observed data for all methods. What is more, Figure 8 display the bivariate plot of duration against severity for observed and simulated data generated from the estimated archimedean copula distribution for six selected stations. Figure 8 indicates that the Gumbel method shows better consistent with the observed data. Therefore, the Gumbel copula was adopted in this paper to calculate the interval probability and return period.

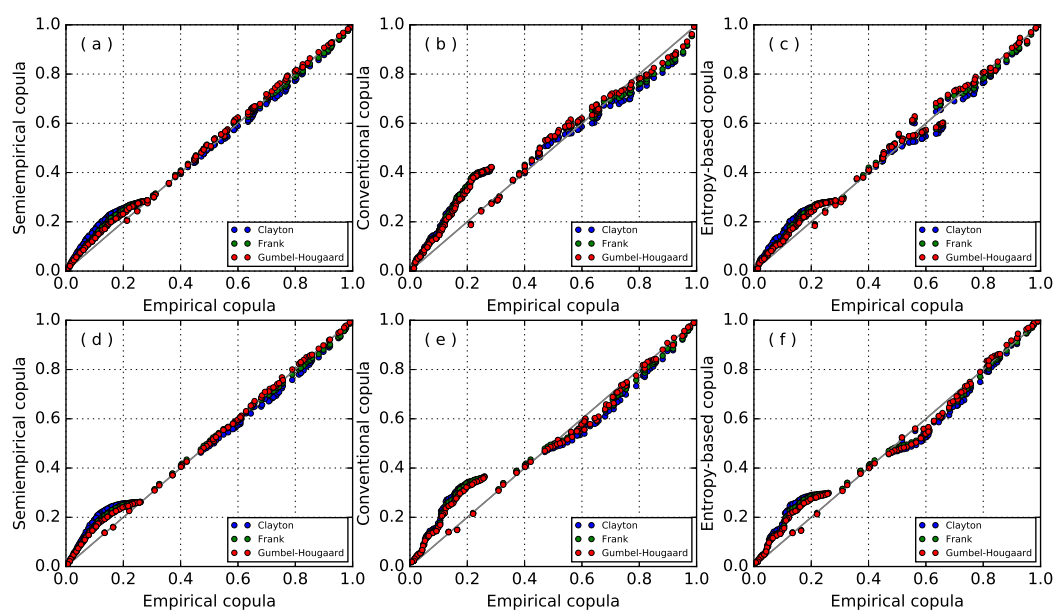


Figure 7. The fitted Archimedean copula versus the empirical copula for Changchun (station No. 54161) and Maerkang (station No. 56172): (a) semiempirical copula for Changchun; (b) conventional copula for Changchun; (c) entropy-based copula for Changchun; (d–f) with same meaning as in subfigures (a–c) but for Maerkang.

Table 3. Percentages of stations in selecting optimal copula function based on *RMSE*, *AIC*, and KS D_n statistic.

Copula Family	Semiempirical			Conventional			Entropy-Based		
	<i>RMSE</i>	<i>AIC</i>	D_n	<i>RMSE</i>	<i>AIC</i>	D_n	<i>RMSE</i>	<i>AIC</i>	D_n
Clayton	0%	0%	0%	0%	0%	9.9%	0%	0%	1.2%
Frank	0.6%	0.6%	0%	0%	0%	0%	1.9%	1.9%	3.7%
Gumbel	99.4%	99.4%	100%	100%	100%	90.1%	98.1%	98.1%	95.1%

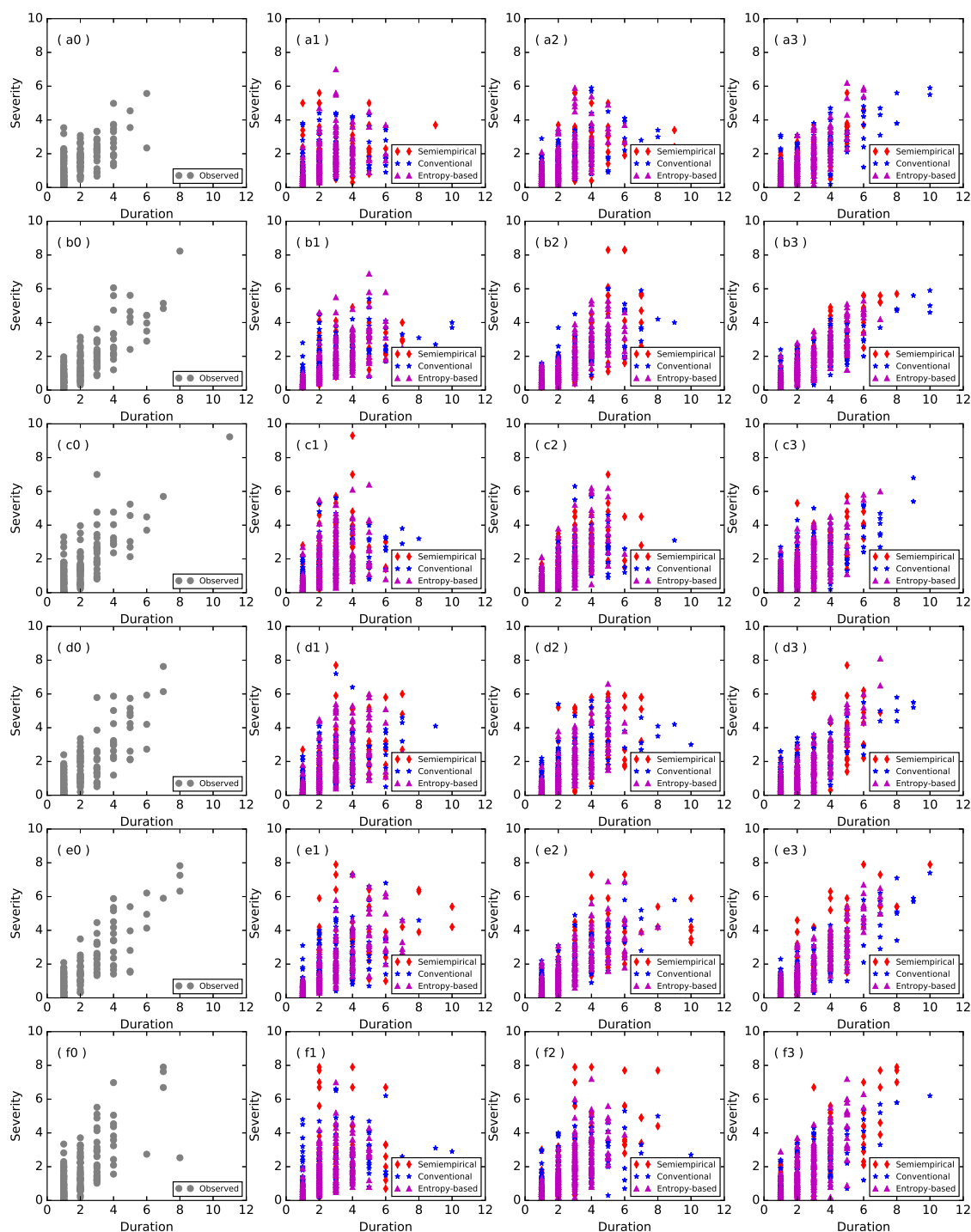


Figure 8. Copula based joint distribution for six stations: (a0–f0) Observed; (a1–f1) Claython; (a2–f2) Frank; (a3–f3) Gumbel; (a–f) as in Figure 5.

4.2. Interval Probability

Conventional processes of bivariate copula analysis usually adopt exponential and gamma distributions to fit drought duration and severity, respectively. Then, from the copula function, the correspondence between drought duration and severity is obtained and the joint probability and return time is calculated. As shown in the literature [17–25], conventional probability analysis focuses on joint cumulative probability and conditional probability. The interval probability was the main

focus of the present study, since it provides a better understanding of drought details and is also suitable for comparing conventional and entropy methods. Therefore the interval probability was obtained as shown in Figures 9–11 for the semiempirical, conventional and entropy-based marginal distributions of the two drought variables (Equations (2), (3), (11), (12), (22) and (23)).

As discussed in Section 3.2, drought events were divided into 16 types (e.g., Figure 9a shows the drought type where $0 < D \leq 1$ and $0 < S \leq 0.85$, referred to as drought type (a), (b), etc. in the following). To compare the probability of each drought type, the spatial probability mean (PM) for each drought type was calculated (shown in the upper center of each subfigure). Because of the small number of stations in western China (Figure 1), the PM was based on data from stations east of longitude 100° E.

As Figures 9–11 show, the PMs of drought types (a), (e), (f) and (k) are obviously larger than the others. The sum of the PMs for these drought types in Figure 9 is equal to 0.774 (0.724 in Figure 9 and 0.781 in Figure 11), indicating that these are the most common types of drought experienced in China. Conversely, the PMs of drought types (c), (d), (h), (i), (m), (n) and (o) approach zero, indicating improbable drought types. In Figures 9–11, types (a), (e) and (f) exhibit significantly different PMs for the three methods. The semiempirical method gives a PM for type (f) which is obviously larger than that of type (a), consistent with the entropy results in Figure 11, whereas Figure 10 shows the opposite. Meanwhile, the PM values for these three drought types indicate that the entropy results are closest to the semiempirical results. The conventional method overestimates the probability of type (a) and underestimates the probability of types (e) and (f). For the extreme drought type (p), the entropy method results match the semiempirical values more closely, but the conventional result is obviously larger.

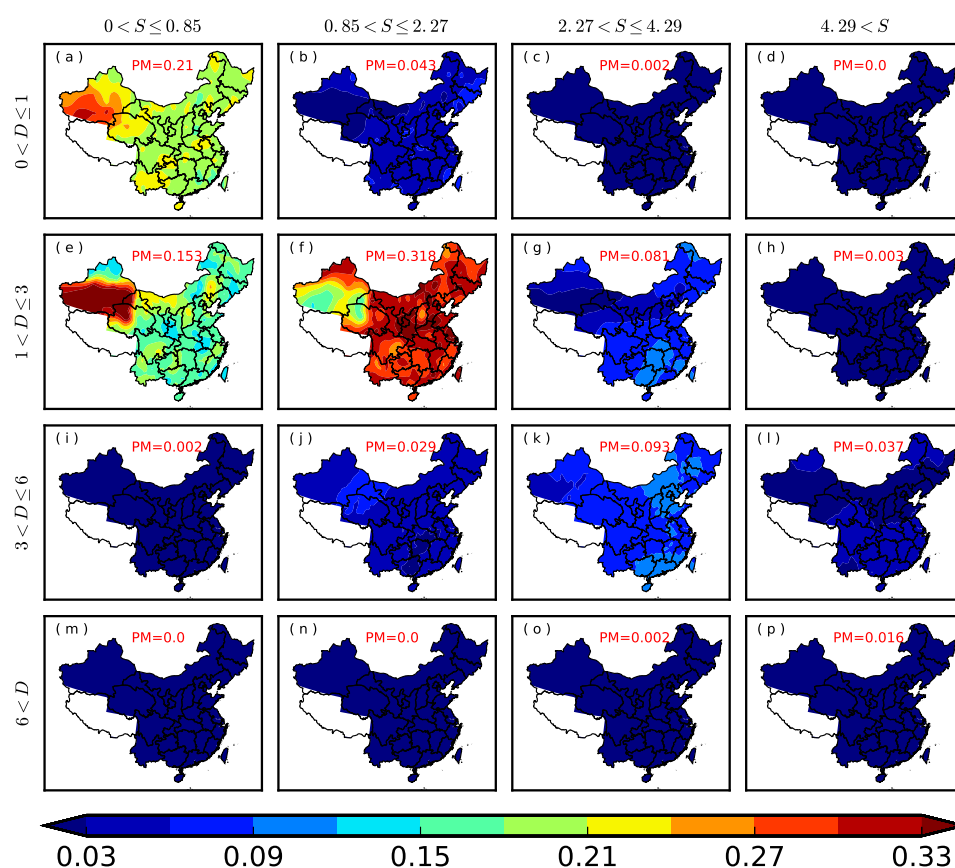


Figure 9. Interval probability based on the semiempirical method: (a–p) represent different drought types.

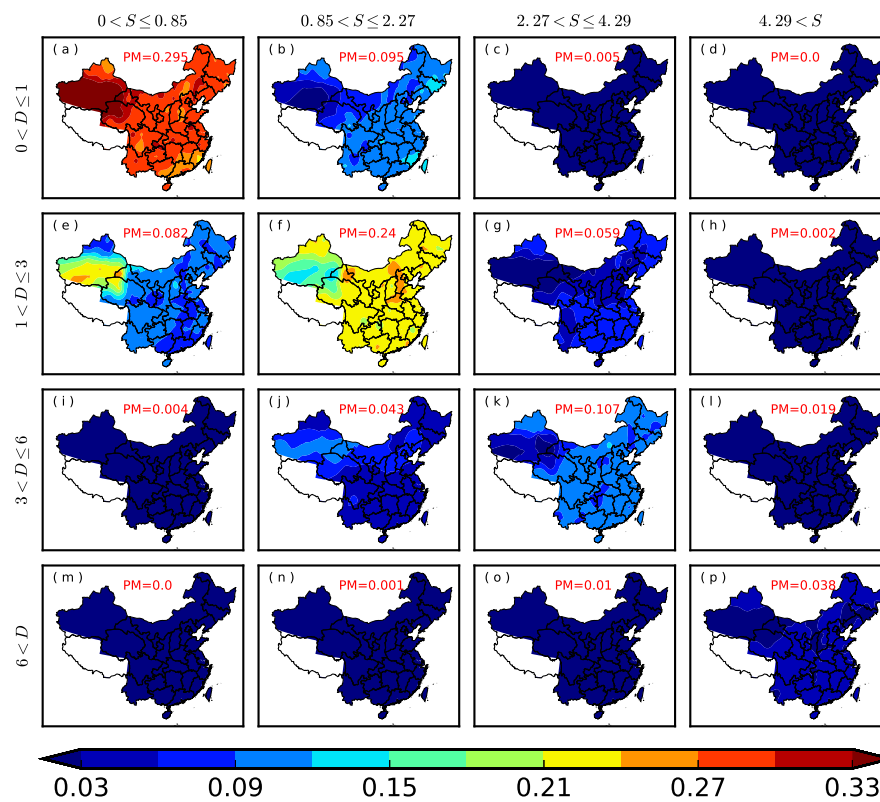


Figure 10. Interval probability based on conventional methods: (a–p) represent different drought types.

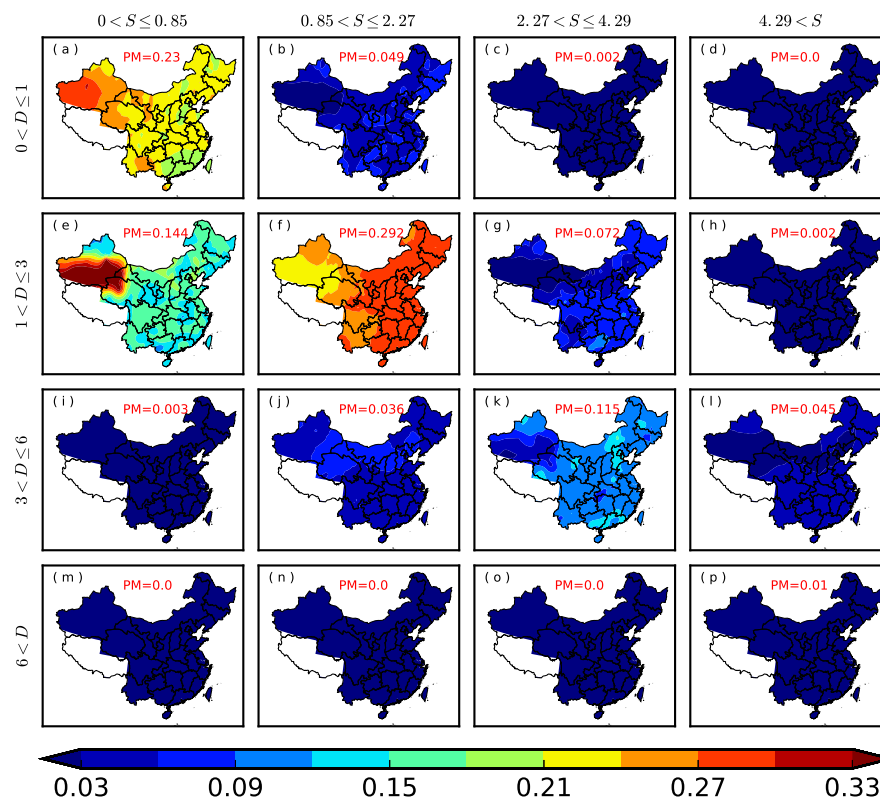


Figure 11. Interval probability based on the entropy method: (a–p) represent different drought types.

4.3. Return Period

As shown in Shiau [17], the joint return period of the bivariate drought event may be calculated from the marginal distributions, the joint probability and the expected drought interarrival time $E(L)$ derivable from the SPI dataset. Calculation of the joint return period can be expressed by either (1) assuming that the drought duration and severity are both larger than the set of values defining the return period $T(D \geq d \text{ and } S \geq s)$; or (2) assuming that one of the drought variables exceeds the set of values defining the return period $T(D \geq d \text{ or } S \geq s)$. However, since the major concern here is the type of drought, the return period (recurrence time for drought type) was calculated from Equation (24). As in the calculation of PM, the mean return period (MRP) was also calculated (Figures 12–14).

As Figures 12–14 show, the return period of drought types (a), (e), (f) and (k) is clearly shorter than for the other types. The MRP illustrates that these drought types would recur every 1–4 years. The spatial variations for these types are also small. Spatial difference is obvious in drought types (l) and (p). As shown for type (l), the return period is less in southern and north-eastern China than in northern China. Drought type (p) in Figure 12 shows a significant spatial discrepancy: in particular, the return period is longer than 1000 years east of $100^\circ E$, the main reason being that these stations have never experienced a type (p) drought, resulting in the semiempirical interval probability $\bar{P}_{d_1, d_2, s_1, s_2}$ equal to 0. Hence the return period cannot be calculated from Equation (24). In practice, a constant return period of 1500 years was allocated to these stations. Therefore, abnormal points merely indicate a longer return period. Such stations were omitted from the calculation of MRP. Hence, based on the MRP, type (p) based on the entropy method was closer to the semiempirical result, whereas the conventional method undervalued the MRP.

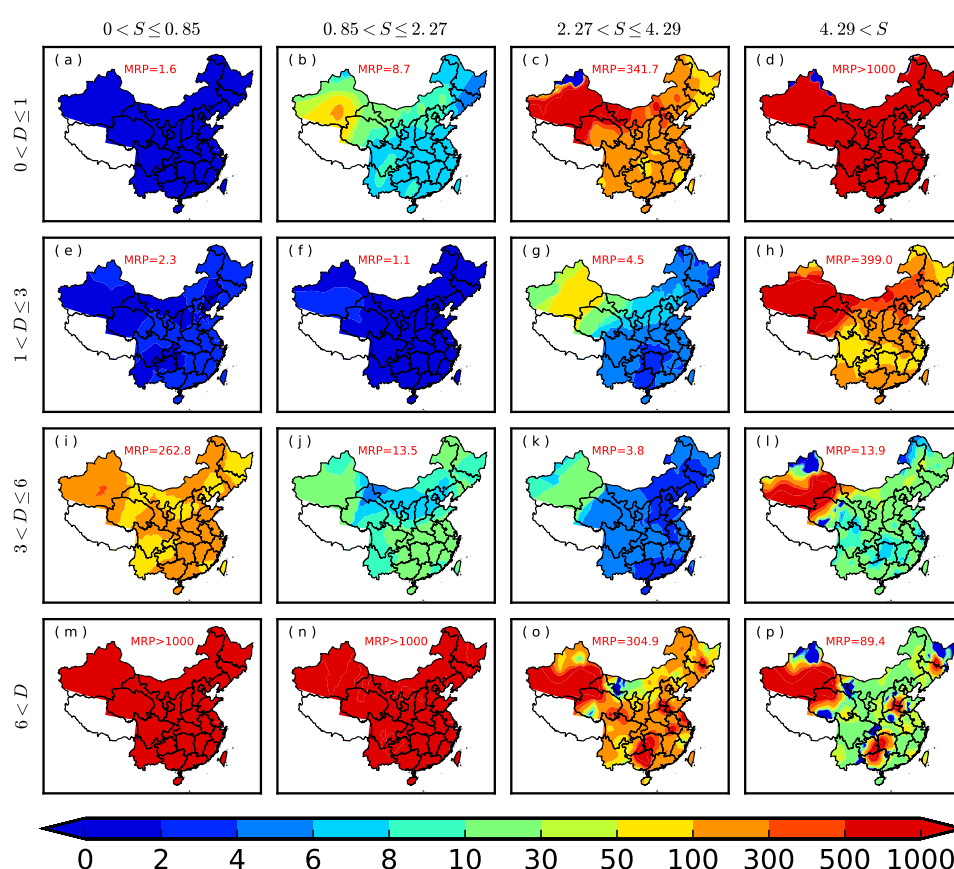


Figure 12. Return period based on semiempirical method (the unit of the colored bar is years). (a–p) represent the different drought types.

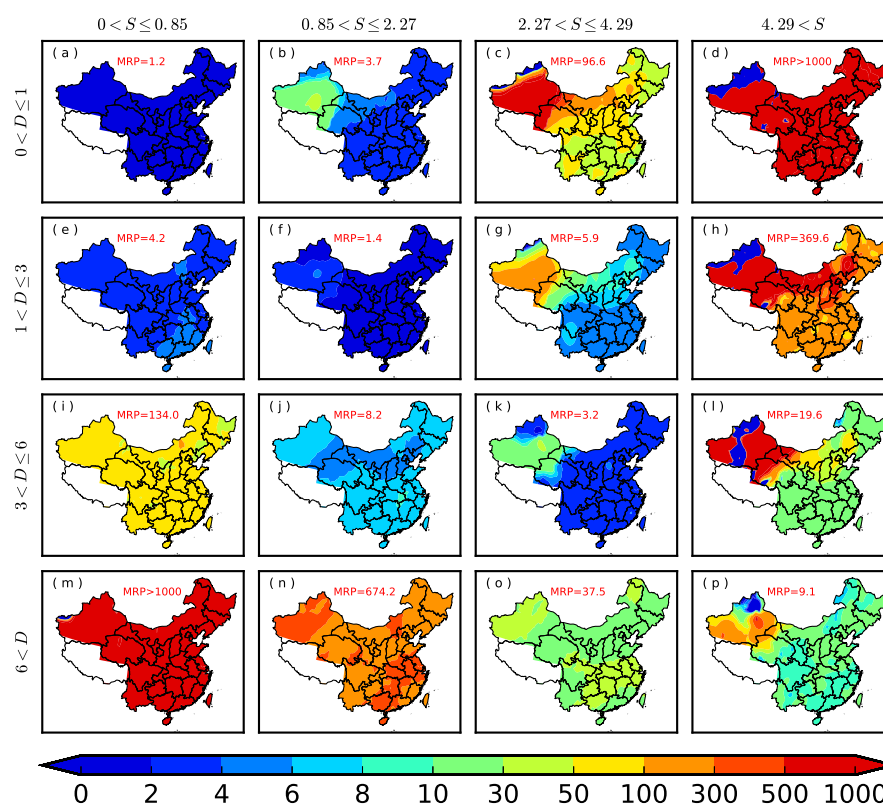


Figure 13. Return period based on conventional method (the unit of the colored bar is years). (a–p) represent the different drought types.

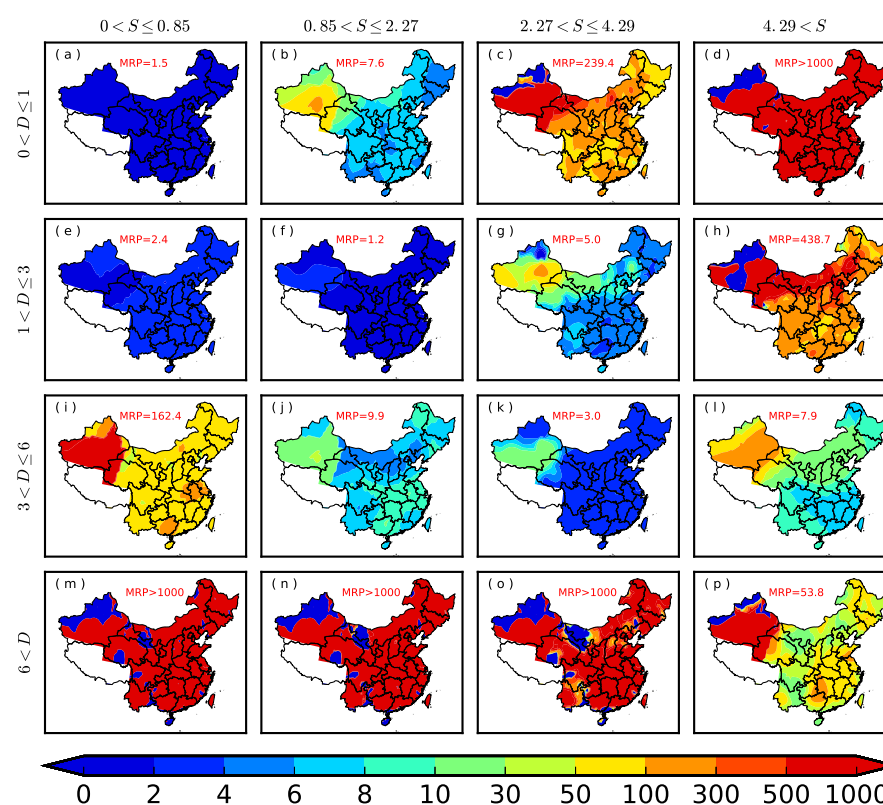


Figure 14. Return period based on entropy method (the unit of the colored bar is years). (a–p) represent the different drought types.

5. Discussion and Conclusions

The entropy distribution is an alternative way of deriving the marginal probability distribution. When the probability distribution of a variable is not known beforehand, the variable with the maximum entropy (subject to certain constraints) is selected to be the distribution function. Because the entropy distribution is unique, it eliminates altogether the need to compare different theoretical distributions. The entropy distribution also fits the data well, as discussed. Since the constraints comprise specified functions with respect to the properties of interest [32], the entropy distribution method is more flexible than conventional methods, which are fixed modeling processes. Thus, we propose the entropy distribution method as an attractive way of obtaining the marginal distribution; however the method has some weaknesses. Firstly, the number of constraints is not fixed. When the number of constraints is small, the fit has a comparatively large deviation, and a large number of constraints results in obvious overfitting—that is, the entropy distribution creates a perfect fit, but only for this dataset. Secondly, both tails of the entropy distribution are unreasonable. Taking drought duration (Equation (11)) as an example, in the left-hand tail of the distribution the probability does not equal 0 when the drought duration is 0, whereas the theoretical distribution (Equation (2)) equals 0, which is the rational value. At the right-hand tail, it is also unreasonable that the Lagrange multiplier λ_3 is negative at some stations, implying a probability greater than 1 when the drought duration is sufficiently large; by contrast, the probability derived from the theoretical distribution is always less than 1. Hence, outside the data range, the probability value given by the entropy distribution is questionable.

This study used the monthly precipitation data recorded at 162 stations for 1961–2015 to compare the conventional method and entropy method, along with the results based on the semiempirical method as the standard of reference. Analysis indicated that the entropy distribution produced a better fit to the data than the theoretical method for the drought duration distribution, although no obvious difference was seen in fitting drought severity. The interval probability and return period analysis showed that drought types (a), (e), (f) and (k) are obviously larger than the other types, and are also the main drought types in China. The conventional and entropy methods gave very different results for types (a) and (f). The entropy method results were closer to the semiempirical results; the conventional method overestimated the probability of type (a) and underestimated the probability of types (e) and (f). Analysis of the return period showed that drought types (a), (e), (f) and (k) would recur every 1–4 years, and no spatial discrepancy was obvious. For the extreme drought types (l) and (p), the entropy method results also agreed more closely with the semiempirical results, whereas the conventional method underestimated the MRP.

In summary, the entropy distribution method may be adopted as an alternative way of deriving the marginal distributions of drought duration and severity in China. The results indicate that the entropy method was more effective than the conventional method for analysis of the probability of drought intervals and return periods. In addition, the entropy distribution can be used for other problems which need to solve the probability distribution expression.

Acknowledgments: The research is supported by the National Natural Science Foundation of China (Grant Nos. 41675092, 41530531 and 41775078), and the Key Special Scientific Research Fund of the Meteorological Public Welfare Profession of China (Grant No. GYHY201506001).

Author Contributions: Jingguo Hu, Wei Hou, and Dongdong Zuo designed research; Dongdong Zuo performed research; Wei Hou, Jingguo Hu, and Dongdong Zuo analyzed the data; Dongdong Zuo and Wei Hou wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Keyantash, J.; Dracup, J.A. The quantification of drought: An evaluation of drought indices. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1167–1180.
2. Sternberg, T. Regional drought has a global impact. *Nature* **2011**, *472*, 169.
3. Wilhite, D.A. Drought as a natural hazard: Concepts and definitions. *Drought Glob. Assess.* **2000**, *1*, 3–18.
4. Begueria, S.; Vicente-Serrano, S.M.; Angulo-Martinez, M. A multiscalar global drought dataset: The speibase a new gridded product for the analysis of drought variability and impacts. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1351–1354.
5. Vicente-Serrano, S.M.; Begueria, S.; Lopez-Moreno, J.I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Clim.* **2010**, *23*, 1696–1718.
6. Huang, J.; Wang, S.W. The experiments of seasonal prediction using the analogy-dynamical model. *Sci. China* **1992**, *35*, 207–216.
7. Huang, J.; Yi, Y.; Wang, S.; Jifen, C. An analogue-dynamical long-range numerical weather prediction system incorporating historical evolution. *Q. J. R. Meteorol. Soc.* **1993**, *119*, 547–565.
8. Feng, G.L.; Dai, X.G.; Wang, A.H.; Chou, J.F. On numerical predictability in the chaos system. *Acta Phys. Sin.* **2001**, *50*, 606–611.
9. Feng, G.L.; Dong, W.J.; Li, J.P. On temporal evolution of precipitation probability of the Yangtze River delta in the last 50 years. *Chin. Phys.* **2004**, *13*, 1582–1587.
10. Zheng, Z.H.; Ren, H.L.; Huang, J.P. Analogue correction of errors based on seasonal climatic predictable components and numerical experiments. *Acta Phys. Sin.* **2009**, *10*, 7359–7367. (In Chinese)
11. Li, J.P.; Ding, R.Q. Temporal-spatial distribution of the predictability limit of monthly sea surface temperature in the global oceans. *Int. J. Climatol.* **2013**, *33*, 1936–1947.
12. Li, J.P.; Wang, S. Some mathematical and numerical issues in geophysical fluid dynamics and climate dynamics. *Commun. Comput. Phys.* **2008**, *3*, 759–793.
13. Sheffield, J.; Wood, E.F. Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *J. Geophys. Res.* **2007**, *112*, D17115.
14. Sheffield, J.; Andreadis, K.M.; Wood, E.F.; Lettenmaier, D.P. Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. *J. Clim.* **2009**, *22*, 1962–1981.
15. Joe, H. Multivariate models and dependence concepts. In *Monographs on Statistics and Applied Probability*; Chapman and Hall: New York, NY, USA, 1997.
16. Nelsen, R.B. *An Introduction to Copulas*; Springer: New York, NY, USA, 1999.
17. Shiau, J.T. Fitting drought duration and severity with two-dimensional copulas. *Water Resour. Manag.* **2006**, *20*, 795–815.
18. Zhang, D.D.; Yan, D.H.; Lu, F.; Wang, Y.C.; Feng, J. Copula-based risk assessment of drought in Yunnan province, China. *Nat. Hazards* **2015**, *75*, 2199–2220.
19. Tosunoglu, F.; Can, I. Application of copulas for regional bivariate frequency analysis of meteorological droughts in Turkey. *Nat. Hazards* **2016**, *82*, 1457–1477.
20. Shiau, J.T.; Shen, H.W. Recurrence analysis of hydrologic droughts of differing severity. *J. Water Res. Plan. Manag.* **2001**, *127*, 30–40.
21. Salas, J.D.; Fu, C.J.; Cancelliere, A.; Dustin, D.; Bode, D.; Pineda, A.; Vencent, E. Characterizing the severity and risk of drought in the Poudre River, Colorado. *J. Water Res. Plan. Manag.* **2005**, *131*, 383–393.
22. Zhang, Q.; Li, J.F.; Singh, V.P. Application of Archimedean copulas in the analysis of the precipitation extremes: Effects of precipitation changes. *Theor. Appl. Climatol.* **2012**, *107*, 255–264.
23. Zhang, Q.; Xiao, M.Z.; Singh, V.P.; Chen, X.H. Copula-based risk evaluation of droughts across the Pearl River basin, China. *Theor. Appl. Climatol.* **2013**, *111*, 119–131.
24. Reddy, M.J.; Singh, V.P. Multivariate modeling of droughts using copulas and meta-heuristic methods. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 475–489.
25. Ganguli, P.; Reddy, M.J. Risk assessment of drought in Gujarat using bivariate copula. *Water Resour. Manag.* **2012**, *26*, 3301–3327.
26. Jaynes, E.T. Information theory and statistical mechanics, I. *Phys. Rev.* **1957**, *106*, 620–630.
27. Jaynes, E.T. Information theory and statistical mechanics, II. *Phys. Rev.* **1957**, *108*, 171–190.

28. Shannon, C.E. A mathematical theory of communications. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
29. Hong, X.; Guo, S.; Xiong, L.; Liu, Z. Spatial and temporal analysis of drought using entropy-based standardized precipitation index: A case study in Poyang Lake basin, China. *Theor. Appl. Climatol.* **2015**, *122*, 543–556.
30. Zhang, L.; Singh, V.P. Bivariate rainfall and runoff analysis using entropy and copula theories. *Entropy* **2012**, *14*, 1784–1812.
31. Hao, Z.; Acse, M.; Singh, V.P.; Asce, F. Entropy-based method for bivariate drought analysis. *J. Hydrol. Eng.* **2013**, *18*, 780–786.
32. Hao, Z.; Singh, V.P. Integrating entropy and copula theories for hydrologic modeling and analysis. *Entropy* **2015**, *17*, 2253–2280.
33. Li, F.; Zheng, Q. Probabilistic modelling of flood events using the entropy copula. *Adv. Water Resour.* **2016**, *97*, 233–240.
34. Mckee, T.B.; Doesken, N.J.; Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the Eighth Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993; American Meteor Society: Boston, MA, USA, 1993; pp. 179–184.
35. Guttman, N.B. Accepting the standardized precipitation index: A calculation algorithm. *J. Am. Water Resour.* **1999**, *35*, 311–322.
36. Hayes, M.J.; Svoboda, M.D.; Wilhite, D.A.; Vanyarkho, O.V. Monitoring the 1996 drought using the standardized precipitation index. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 429–438.
37. Bordi, I.; Fraedrich, K.; Jiang, J.M.; Suter, A. Spatio-temporal variability of dry and wet periods in eastern China. *Theor. Appl. Climatol.* **2004**, *79*, 81–91.
38. Livada, I.; Assimakopoulos, V.D. Spatial and temporal analysis of drought in Greece using the Standardized Precipitation Index (SPI). *Theor. Appl. Climatol.* **2007**, *89*, 143–153.
39. Zhang, Q.; Xu, C.Y.; Zhang, Z. Observed changes of drought/wetness episodes in the Pearl River basin, China, using the standardized precipitation index and aridity index. *Theor. Appl. Climatol.* **2009**, *98*, 89–99.
40. Mathier, L.; Perreault, L.; Bobée, B.; Ashkar, F. The use of geometric and gamma-related distributions for frequency analysis of water deficit. *Stoch. Hydrol. Hydraul.* **1992**, *6*, 239–254.
41. Kapur, J.N. *Maximum-Entropy Models in Science and Engineering*; John Wiley & Sons Inc.: New York, NY, USA, 1989.
42. Kesvan, H.; Kapur, J. *Entropy Optimization Principles with Applications*; Academic Press: New York, NY, USA, 1992.
43. Mead, L.R.; Papanicolaou, N. Maximum entropy in the problem of moments. *J. Math. Phys.* **1984**, *8*, 2404–2417.
44. Sklar, M. Fonctions de repartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
45. Mirabbasi, R.; Fakheri-Fard, A.; Dinpashoh, Y. Bivariate drought frequency analysis using the copula method. *Theor. Appl. Climatol.* **2012**, *108*, 191–206.
46. Akaike, H. IEEE Xplore abstract-a new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
47. Zhang, Q.; Singh, V.P. Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.* **2006**, *11*, 150–164.
48. Rauf, U.F.A.; Zeephongsekul, P. Copula based analysis of rainfall severity and duration: A case study. *Theor. Appl. Climatol.* **2014**, *115*, 153–166.
49. FruH, B.; Feldmann, H.; Panitz, H.J.; Schadler, G. Determination of Precipitation Return Values in Complex Terrain and Their Evaluation. *J. Clim.* **2010**, *23*, 2257–2274.
50. Gao, H.; Yang, S. A severe drought event in northern China in winter 2008–2009 and the possible influences of La Niña and Tibetan Plateau. *J. Geophys. Res.* **2009**, *114*, D24104.
51. Qian, W.; Shan, X.; Zhu, Y. Ranking regional drought events in China for 1960–2009. *Adv. Atmos. Sci.* **2011**, *28*, 310–321.
52. Xin, X.G.; Yu, R.C.; Zhou, T.J.; Wang, B. Drought in late spring of south China in recent decades. *J. Clim.* **2006**, *19*, 3197–3206.
53. Xin, X.G.; Yu, R.C.; Zhou, T.J. Southward movement of the decadal drought in southeastern China during April–May and numerical simulation of the effect of the condensation heating. *Chin. J. Atmos. Sci* **2009**, *33*, 1165–1173. (In Chinese)

54. Yu, W.J.; Shao, M.Y.; Ren, M.L.; Zhou, H.J.; Jiang, Z.H.; Li, D.L. Analysis on spatial and temporal characteristics drought of Yunnan Province. *Acta Ecol. Sin.* **2013**, *33*, 317–324.
55. Yang, J.; Gong, D.Y.; Wang, W.S.; Hu, M.; Mao, R. Extreme drought event of 2009/2010 over southwestern China. *Meteorol. Atmos. Phys.* **2012**, *115*, 173–184.
56. Huang, R.H.; Liu, Y.; Wang, L.; Wang, L. Analyses of the causes of severe drought ccurring in southwest China from the fall of 2009 to the spring of 2010. *Chin. J. Atmos. Sci.* **2012**, *36*, 443–457. (In Chinese)



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).